Bayesian two-component measurement error modelling for survival analysis using INLA—A case study on cardiovascular disease mortality in Switzerland

Muff, Stefanie ; Ott, Manuela ; Braun, Julia ; Held, Leonhard

Abstract: Measurement error (ME) in explanatory variables is a common problem in regression and survival analysis, as it may cause bias in the estimated parameters. It is shown how the integrated nested Laplace approximations (INLA) method can handle classical and Berkson ME in a single explanatory variable, illustrated for the case of a Weibull regression model. To this end, a two-component error model to account for a mix of Berkson and classical ME in a single covariate is introduced and applied to a study on cardiovascular disease (CVD) mortality in Switzerland. In particular, the model was used to correct for error in the self-reported mean daily number of cigarettes smoked, as well as in reportings of systolic blood pressure (SBP). Both variables suffer from classical error induced by an imprecision, either due to misremembering of study participants (cigarettes), or due to practical difficulties in obtaining accurate measurements (SBP), but also from a Berkson-type error that is induced by a rounding behaviour, also known as digit preference. In both cases, the effect estimates increased when the error was taken into account. Therefore, an important conclusion is that ME modelling in survival analysis is relevant, and a ready-to-use Bayesian solution including R-code is provided.

DOI: https://doi.org/10.1016/j.csda.2017.03.001

Bayesian measurement error modelling for survival analysis using INLA - A case study on cardiovascular disease mortality in Switzerland

Stefanie Muff[a,b,1,*], Manuela Ott[a,1], Julia Braun[a], Leonhard Held[a]

[a] *Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland*

[b] *Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

Abstract

Measurement error (ME) in explanatory variables is a common problem in regression and survival analysis, as it may cause bias in the estimated parameters. It is shown how the integrated nested Laplace approximations (INLA) method can handle classical and Berkson ME in a single explanatory variable, illustrated for the case of a Weibull regression model. To this end, a two-component error model to account for a mix of Berkson and classical ME in a single covariate is introduced and applied to a study on cardiovascular disease (CVD) mortality in Switzerland. In particular, the model was used to correct for error in the self-reported mean daily number of cigarettes smoked, as well as in reportings of systolic blood pressure (SBP). Both variables suffer from classical error induced by an imprecision, either due to misremembering of study participants (cigarettes), or due to practical difficulties in obtaining accurate measurements (SBP), but also from a Berkson-type error that is induced by a rounding behavior, also known as digit preference. In both cases, the effect estimates increased when the error was taken into account. Therefore, an important conclusion is that ME modelling in survival analysis is relevant, and a ready-to-use Bayesian solution including `R`-code is provided.

Keywords: Measurement error; Weibull regression; Bayesian analysis; cardiovascular disease (CVD); integrated nested Laplace approximations (INLA)

*Corresponding author

Email addresses: `stefanie.muff@uzh.ch` (Stefanie Muff), `manuela.ott@uzh.ch` (Manuela Ott), `julia.braun-gruebel@uzh.ch` (Julia Braun), `leonhard.held@uzh.ch` (Leonhard Held)

[1] These authors contributed equally

1. Introduction

Survival or time-to-event data frequently arise in medical applications. When a regression model, such as the Cox or the Weibull model, is used to relate certain covariates to the survival outcome, ME in covariates is often present, but mostly ignored. However, it is well known that ME in covariates of survival models or generalized linear mixed models (GLMMs) may lead to biased estimates of the parameters (Fuller, 1987; Carroll et al., 2006). There is now a large body of literature on methods to model and adjust for ME in covariates including likelihood approaches (Carroll et al., 1984; Schafer, 1987, 1993), score function methods (Stefanski, 1989; Nakamura, 1990), method-of-moments corrections (Fuller, 1987), simulation extrapolation (SIMEX) (Cook and Stefanski, 1994; Küchenhoff et al., 2006), regression calibration (Carroll and Stefanski, 1990; Gleser, 1990), or Bayesian analyses (Lindley and El-Sayyad, 1968; Clayton, 1992; Stephens and Dellaportas, 1992; Richardson and Gilks, 1993; Dellaportas and Stephens, 1995; Gustafson, 2004; Muff et al., 2015).

For survival outcomes, Gimenez et al. (1999) have considered covariate ME in the parametric Weibull model, but most references focus on covariate ME in the Cox regression model (Cox, 1972), see for instance Prentice (1982), Nakamura (1992), Hughes (1993), Hu et al. (1998), Song et al. (2002), Tadesse et al. (2005) and Küchenhoff et al. (2007). While the Bayesian approach to model covariate error is rather popular for GLMMs, errors in survival models have only rarely been addressed with Bayesian methods. An interesting exception is Tadesse et al. (2005), who took a Bayesian approach to ME adjustments in Cox-type regression using a hierarchical model, estimated by Markov chain Monte Carlo (MCMC) methods. If the latent field of the Bayesian hierarchical model is Gaussian, an alternative estimation procedure is INLA (Rue et al., 2009), which is able to handle GLMMs (Fong et al., 2010), but also several types of survival models, including Cox and Weibull regression (Martino et al., 2011). In addition, Muff et al. (2015) have used INLA for ME adjustments in covariates of GLMMs. The advantages of the INLA approach over MCMC methods are the absence of Monte Carlo error in the estimates and, perhaps even more importantly, its computational speed. In fact, Muff et al. (2015) reported computation times that were several orders of magnitudes shorter than those used for MCMC sampling in a standard logistic regression example, where one covariate affected by ME was modelled.

In this paper, we have combined Bayesian ME modelling with Weibull survival analysis using INLA in the context of a cohort study in Switzerland aiming at CVD prevention, where considerable ME in some covariates is expected. The dataset was previously analyzed by Von Gunten et al. (2013) and, as a main result, it was found that high plasma glucose concentrations ($\geq$ 6.1 mmol/L) at baseline results in a 1.06 fold risk increase (with 95% confidence interval from 1.00 to 1.12) to die of CVD over a mean follow-up period of 25 years. However, while the precision of plasma glucose

measurements is not of concern, two other covariates that also led to increased CVD mortality in the study contain considerable ME: the mean number of cigarettes smoked per day and the SBP. Cigarette consumption was assessed by patients' self-reports, which were previously shown to be subject to two types of error: While the actual cigarette consumption of the participants is subject to misremembering, also known as *recall bias* (Wang et al., 2012), most participants tend to report multiples of five, ten or twenty (Klesges et al., 1995; Wang and Heitjan, 2008), a behavior that is known as *rounding*, *heaping* (Zelnik, 1961; Wang and Heitjan, 2008) or *digit preference* (Myers, 1954; Camarda et al., 2008). On the other hand, blood pressure is notoriously difficult to measure, mainly because it is subject to large temporary fluctuations, but also due to imprecision in the devices and their handling (Carroll et al., 2006). Moreover, some end-digit preference in blood pressure recordings has been observed (de Lusignan et al., 2004), hinting at additional rounding errors.

Adjusting for error in these two covariates is important and may have an impact on the findings of Von Gunten et al. (2013). We have therefore reanalyzed their data, adding ME models for both the mean number of cigarettes smoked per day and for the SBP, first separately and then jointly for both covariates. As a main result, error modelling led to larger estimated effects for both covariates, indicating that the two variables might reduce expected survival times more than suggested by the analysis that did not account for error in the covariates.

From a methodological point of view, the analysis of this dataset required three generalizations of the ME models presented in Muff et al. (2015): First, and most importantly, we introduced a two-component error model to account for two different ME mechanisms, called classical and Berkson ME, in the same covariate. Such a two-component error model was applied to the cigarette counts as well as to the SBP to model the misremembering/mismeasuring and the rounding errors jointly. In comparison, only simple classical and Berkson error models were introduced previously to be handled with INLA. Second, we fitted models where ME in multiple covariates was modelled jointly. And third, we extended the exponential family likelihoods that were used previously to Weibull survival models that account for censoring. The rationale to select the Weibull model was that Cox-type survival models are usually handled within the Bayesian framework, and in particular in INLA, via the approximation of the baseline hazard by a log-constant version of it (Tadesse et al., 2005; Martino et al., 2011), which requires a careful partitioning of the time axis into intervals, as the results may depend on this choice. All analyses were carried out using the `R`-interface `R-INLA`, which can be downloaded from `http://www.r-inla.org`.

This paper is organized as follows: In Section 2, we introduce the dataset on CVD and all-cause mortality in Switzerland that was analyzed in Von Gunten et al. (2013). Section 3 describes the Weibull regression model, and two commonly used error models

(Berkson and classical), as well as a two-component error model including a classical and a Berkson component. In Section 4 the layers of the Bayesian hierarchical ME models are listed, and we explain how those can be jointly modelled using the INLA framework. Section 5 first relates the Weibull regression to the Cox regression model used by Von Gunten et al. (2013), and deals with the application of this methodology to the dataset introduced in Section 2. Finally, we conclude with a discussion in Section 6.

2. Survival analysis in the Swiss National Cohort Study

The National Research Program 1A (NRP 1A) in Switzerland (1977–1979) aimed at CVD prevention. 8631 participants (men and women) were enrolled (Gutzwiller et al., 1985). They attended an initial health examination and completed a detailed questionnaire. In the original study concept, no mortality follow-up was planned. However, information on mortality status and - if applicable - cause of death could be obtained by anonymous record linkage with the Swiss National Cohort (SNC). As explained in Bopp et al. (2012), the SNC includes information on all residents of Switzerland from the Swiss national census in the years 1990 and 2000, as well as information on death or emigration. From the information on 8631 original participants, data of 8008 persons could be linked with follow-up survival information from the SNC up to the year 2008.

Von Gunten et al. (2013) fitted Cox regression models to the linked NRP 1A dataset for the outcomes all-cause and CVD mortality, stratified for men and women. 24 persons had to be excluded due to missing plasma glucose measurements or because they were younger than 16 years, leading to a study population of 7984 persons. The main explanatory variable was blood plasma glucose concentration (mmol/L), and the following predictors were included as additional covariates in the analysis: age at study entry (years), time since last meal (fasting time, h), marital status (single, married, widowed and separated or divorced), town of residence (Aarau, Solothurn, Nyon, Vevey and Lugano), physical activity score (0-8), a binary indicator for three main meals per day (a proxy for healthy eating habits), body mass index (BMI, kg/m^2), serum cholesterol (mmol/L), SBP (mm Hg, mean of up to four measurements), and the average number of cigarettes smoked per day (self-reports from the participants). Von Gunten et al. (2013) found that high plasma glucose concentrations ($\geq$ 6.1 mmol/L) were associated with increased risk of CVD and all-cause mortality, and that age, cholesterol concentration, SBP and smoking also had an impact on CVD mortality.

Here, we will focus on CVD mortality of the $n = 3607$ male participants in the NRP 1A study. To understand the effect of ME modelling in covariates, we will reanalyze survival of the participants with respect to CVD mortality using a Weibull regression model and the same covariates as in the original study. Summary statistics of all covariates including the survival time (in days) are given in Tables 1 and 2. The

Variable	**Min**	$\mathbf{q_1}$	$\bar{\mathbf{x}}$	$\mathbf{q_3}$	**Max**	**s**
Survival time [days]	40.0	7204.5	9153.4	11355.0	11444.0	3253.4
Rescaled survival time	0.0	0.6	0.8	1.0	1.0	0.3
Glucose [mmol/L]	1.8	4.9	5.5	5.8	24.6	1.2
Age [years]	16.1	30.7	42.1	53.1	89.6	15.1
No. of cigarettes per day	0.0	0.0	6.5	10.0	80.0	11.1
Mean SBP [mm Hg]	84.0	120.0	130.0	138.0	265.0	16.7
Mean log(SBP-50)	3.5	4.2	4.4	4.5	5.4	0.2
Cholesterol [mmol/L]	5.2	5.4	5.9	6.3	7.8	0.7
BMI [kg/m^2]	14.2	22.5	24.8	26.8	41.5	3.3
Fasting time [h]	0.0	1.9	4.4	5.5	23.9	3.8
Physical activity score [0-8]	0.0	3.0	3.8	5.0	8.0	1.7

Table 1: Summary statistics for the survival times and the continuous covariates for all $n = 3607$ individuals included in the analyses. Since we will use a log-transformed version of the mean SBP in all the following models, the distribution of the transformed SBP is also shown here. The table shows (from left to right) the minimum, lower quantile, mean, upper quantile, maximum value and the standard deviation for each covariate.

survival time of a participant was censored if he was still alive at the end of the follow-up period, or if he had died from a cause different than CVD in the observation period. Among the considered 3607 participants with sufficient information on survival, 458 deaths attributed to CVD were observed in the follow-up period of 32 years. 87.3% of the observations are thus censored.

Variable	**Levels**	**n**	%
3 main meals	No	1234	34.2
	Yes	2373	65.8
	all	3607	100.0
Marital status	Married	2668	74.0
	Widowed	50	1.4
	Divorced or separated	127	3.5
	Single	762	21.1
	all	3607	100.0
Town	Aarau	1332	36.9
	Lugano	198	5.5
	Nyon	1127	31.2
	Solothurn	539	14.9
	Vevey	411	11.4
	all	3607	100.0

Table 2: Summary statistics for the categorical covariates.

Note that the SBP covariate was substituted here by $\log(\text{SBP} - 50)$, a transformation that helps to achieve approximate normality for the set of observations, which is important when modelling the observed blood pressure as a proxy for the true covariate in the INLA framework. The same transformation has been used previously to model ME in the measurements of SBP (Carroll et al., 2006; Muff et al., 2015). Repeated SBP measurements were available for a subset of patients, and the average of the (untransformed) SBP repeats was used as covariate in the original analysis. While 75% of the participants had one SBP measurement only, 20.9% had two measurements

and the remaining 4.1% had three or four measurements (see Table S1 in the Supplementary Material). The first SBP measurement was available for all participants. If a participant had this first measurement $\geq$ 140 mm Hg or if his measured diastolic blood pressure was $\geq$ 90 mm Hg, a second SBP measurement was taken during the same examination (study protocol), which resulted in a dependency among the first two measurements. In fact, the empirical correlation between the first and the second SBP measurements $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ was higher (≈ 0.79) than the correlation between any other pair (j, k) of repeated measurements (values between 0.58 and 0.64). Given that such a study protocol results in biased measurements, the second measurement was discarded from all analyses described here, that is, only measurements 1, 3 and 4 were used throughout the paper if they were available. In addition, four participants had the third and one of them also the fourth SBP measurement smaller than 20 mm Hg, and these implausible values were replaced by missing values. The average of the remaining transformed repeats was then included as a covariate.

3. Weibull regression and error models

In this section we will first describe the Weibull regression model and the parameterization that was used here. Two most commonly used error models for continuous covariates, namely the Berkson and the classical error models, as well as a two-component error model with a mixture of Berkson and classical error, will then be formulated for later use in the application to the NRP 1A study data.

3.1. Weibull model

The Weibull survival regression model is a fully parametric alternative to Cox regression. Both the Weibull and the Cox model assume proportional hazards, and both lead to asymptotically unbiased estimates of the hazard ratio (Carroll, 2003). An unquestionable advantage of the Cox model is that no assumptions about the underlying baseline hazard are made (Cox, 1972; Kalbfleisch and Prentice, 2002; Collett, 2015). On the other hand, Weibull regression provides the only proportional hazards model that also leads to an accelerated failure-time (AFT) model (Prentice and Kalbfleisch, 1979; Carroll, 2003).

Denote by $\tilde{t}_i$ the survival time and c_i the censoring time of the i-th participant. Then, the observed time is given by $t_i = \min\{\tilde{t}_i, c_i\}$ with an indicator δ_i for CVD death, defined as

$$\delta_i = \begin{cases} 1 & \text{if participant } i \text{ died from CVD at time } t_i \\ 0 & \text{otherwise .} \end{cases},$$

Throughout the paper, we rescaled the observed survival times to the interval $[0, 1]$, with 1 corresponding to the longest observed survival time (*i. e.* 11 444 days). We fitted Weibull survival models to $(\boldsymbol{t}, \boldsymbol{\delta})$, where $\boldsymbol{t} = (t_1, \ldots, t_n)^\top$ is the vector of rescaled observed survival times and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^\top$ the vector of censoring indicators for death due to CVD. Given an error-prone covariate vector $\boldsymbol{x}$ whose actual values are unobserved, a matrix of error-free covariates $\boldsymbol{z}$ with covariate vectors as columns and measurements $\boldsymbol{z}_i$ of individual i as rows, and (vectors of) regression coefficients β_x and $\boldsymbol{\beta}_z$, the linear predictor for the i-th individual can be written as

$$\eta_i = \beta_0 + x_i \beta_x + \boldsymbol{z}_i^\top \boldsymbol{\beta}_z \ . \tag{1}$$

The extension of the equation to several error-prone covariates is straightforward, but we avoid a more general notation for the moment to keep notation simple. The Weibull distribution was parameterized with shape parameter γ and scale parameter $\lambda_i = \exp(\eta_i)$. The hazard to die at time $t \in [0, 1]$ after study entry for the i-th individual is then given as

$$h_i(t) = \exp(\eta_i)\, \gamma\, t^{\gamma - 1} \ , \tag{2}$$

see *e. g.* Collett (2015). Thanks to the interpretation of the Weibull model as an AFT model, estimates of the effects on expected survival times from changing covariate values can be obtained from a simple formula (see equation (3) below). The parametric assumption of the Weibull model therefore pays off in that the results do not only provide information on the event rates (*i. e.* the hazard ratio), but also on relative changes in survival times, which can be assessed via the calculation of an *acceleration factor*, denoted as the event time ratio (ETR). Using our parameterization of the Weibull regression model, the ETR for the covariate in the k-th column, $k \geq 2$, of the matrix $\boldsymbol{v} = \begin{bmatrix} \mathbf{1} & \boldsymbol{x} & \boldsymbol{z} \end{bmatrix}$ is given by

$$\kappa^{(k)} = \exp(-\beta^{(k)}/\gamma) \ , \tag{3}$$

where $\beta^{(k)}$ is the k-th component of the vector $\boldsymbol{\beta} = (\beta_0, \beta_x, \boldsymbol{\beta}_z^\top)^\top$ ($k = 1$ corresponds to the intercept, in which we are not interested here). This ratio quantifies the proportional change expected in survival times resulting from a change by one unit in a continuous covariate, or between two factor levels (Carroll, 2003). More generally, the relative change in survival times between individuals i and j with covariate vectors $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ is given by $\exp(-\boldsymbol{\beta}^\top(\boldsymbol{v}_j - \boldsymbol{v}_i)/\gamma)$.

3.2. Berkson measurement error

Berkson-type error in covariates of regression models typically occurs in experimental setups (*e. g.* due to deviations from predefined, planned doses or concentrations),

epidemiological studies (*e. g.* when assigning exposure measurements to individuals within a radius of a measurement station), or when a rounded proxy $\boldsymbol{w}$ of a true variable $\boldsymbol{x}$ is used (*e. g.* due to limited precision of a measurement device). Examples are the application of fixed doses of herbicides in bioassay experiments (Rudemo et al., 1989) or the radiation epidemiology study described in Kerber et al. (1993) and Simon et al. (1995). Let the vector $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ be the proxy for an unobserved (latent) covariate $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$, and $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$ the vector of corresponding error terms. The Berkson measurement error model (Berkson, 1950) is then given as

$$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{u} \ , \quad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \tau_u \boldsymbol{D}) \ , \tag{4}$$

where $\boldsymbol{w}$ and $\boldsymbol{u}$ are assumed independent given the other covariates $\boldsymbol{z}$. The diagonal matrix $\boldsymbol{D}$ contains entries $d_i > 0$ to account for heteroscedasticity, with $\boldsymbol{D}$ equal to the identity matrix if the error model is homoscedastic, and error precision τ_u. Please note that we parameterize the multivariate normal distribution with mean and precision matrix, where the precision matrix is the inverse of the covariance matrix. One important characteristic of Berkson ME is that, due to independence of $\boldsymbol{w}$ and $\boldsymbol{u}$, $\mathrm{Var}(\boldsymbol{x}) = \mathrm{Var}(\boldsymbol{w}) + \mathrm{Var}(\boldsymbol{u})$, *i. e.* the variance of the true unobserved covariate $\boldsymbol{x}$ is larger than the variance of the proxy $\boldsymbol{w}$. Finally, we assume that the survival outcome $(\boldsymbol{t}, \boldsymbol{\delta})$ is conditionally independent of the proxy $\boldsymbol{w}$, given the true covariates $\boldsymbol{x}$ and $\boldsymbol{z}$. Such an ME model is called non-differential (Carroll et al., 2006) and implies that the proxy $\boldsymbol{w}$ contains no additional information on the response $(\boldsymbol{t}, \boldsymbol{\delta})$ if the true covariates $\boldsymbol{x}$ and $\boldsymbol{z}$ are known.

3.3. Classical measurement error

While Berkson error mainly occurs in experimental setups, the classical-type error model can usually be found in the context of measurements, which may be taken repeatedly, *e. g.* in the field or in the laboratory. A prominent example are SBP values of study participants (Carroll et al., 2006), which were used as a covariate in the NRP 1A study investigated here. The observed proxy $\boldsymbol{w}$ is then a composition of the true covariate $\boldsymbol{x}$ and the error term $\boldsymbol{u}$, *i. e.*

$$\boldsymbol{w} = \boldsymbol{x} + \boldsymbol{u} \ , \quad \boldsymbol{u} \sim \mathrm{N}(\boldsymbol{0}, \tau_u \boldsymbol{D}) \ , \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{u}$ are assumed independent given $\boldsymbol{z}$, and $\boldsymbol{D}$ is a diagonal matrix with positive entries. Ideally, repeated measurements w_{ij}, $j = 1, \ldots, J_{n_i}$, of the true value are available, so that $w_{ij} \,|\, x_i \sim \mathrm{N}(x_i, \tau_u d_i)$. The repeated measurements w_{ij} are sometimes unbalanced, but usually assumed to be conditionally independent. In contrast to the Berkson ME, the variance of the observed proxy $\boldsymbol{w}$ in the presence of classical ME is

larger than the variance of the true covariate, *i. e.* $\mathrm{Var}(\boldsymbol{w}) > \mathrm{Var}(\boldsymbol{x})$. We again assume a non-differential ME model.

3.4. Two-component error model

A true covariate is sometimes blurred by multiple mechanisms, which may induce classical and Berkson error components in the same covariate (Heid et al., 2004). For instance, assume that an imprecisely measured covariate is rounded to a convenient number. In this case, classical error (for the imprecision of the measurement) and Berkson error (for the rounding) are present in the same variable. Note that this is the situation we encounter for the cigarette consumption values that were reported by the smoking individuals in the NRP 1A study: While participants are likely to misremember their actual mean cigarette consumption (Wang et al., 2012), some smokers preferentially reported multiples of 5, 10 or 20 (Figure 1, left). A similar issue arises for blood-pressure recordings, which are difficult to measure (classical error), and in addition have been shown to suffer from an over-proportional representation of end-digits 0, 5 and the even numbers 2, 4, 6 and 8, corresponding to a human-induced rounding (Berkson) error (de Lusignan et al., 2004). The distribution of end-digits in SBP recordings of the present study is shown in the right panel of Figure 1.

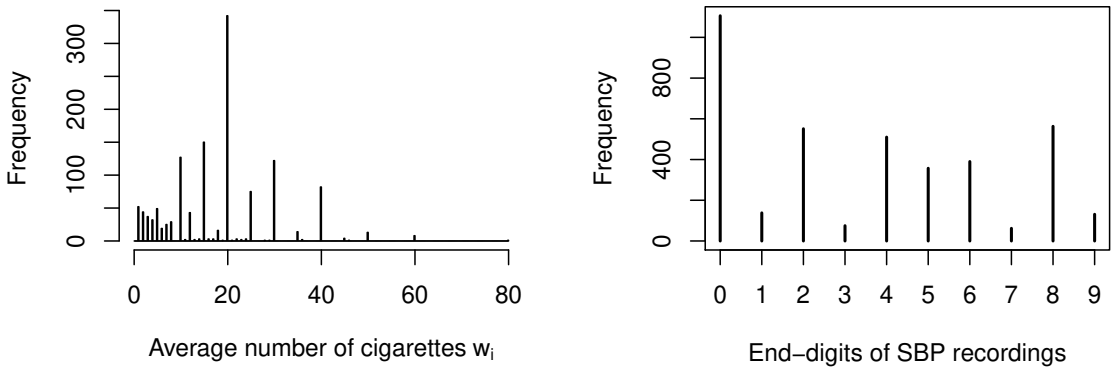

Figure 1: Frequencies of self-reported average number of cigarettes smoked per day, excluding non-smokers, where the peaks at multiples of five and ten are clearly visible (left), and frequencies of observed end-digit recordings for SBP (right).

To model such an error structure, the trick is to include an additional latent variable $\boldsymbol{r}$ for the (unobserved) mismeasured values, that is, for the remembered values by the smokers or for the actual SBP displayed by the measurement device, respectively. The

true values $\boldsymbol{x}$, the reported values $\boldsymbol{w}$ and $\boldsymbol{r}$ are then related as

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{x} + \boldsymbol{u}_c \,, \quad \boldsymbol{u}_c \sim \mathrm{N}(\boldsymbol{0}, \tau_{u_c} \boldsymbol{D}_c) \quad \text{and} & (6) \\ \boldsymbol{r} &= \boldsymbol{w} + \boldsymbol{u}_b \,, \quad \boldsymbol{u}_b \sim \mathrm{N}(\boldsymbol{0}, \tau_{u_b} \boldsymbol{D}_b) \,, & (7) \end{aligned}$$

where $\boldsymbol{u}_c$ and $\boldsymbol{u}_b$ are the classical and Berkson error terms, and τ_{u_c} and τ_{u_b} are the respective error precisions. The first equation relates the latent mismeasured or misremembered values $\boldsymbol{r}$ to the true (but also latent) values $\boldsymbol{x}$, while the second equation implies that the reported values $\boldsymbol{w}$ are rounded versions of $\boldsymbol{r}$. The diagonal matrices $\boldsymbol{D}_c$ and $\boldsymbol{D}_b$ contain as entries for individual i the scaling constants $d_i^{(c)}$ and $d_i^{(b)}$, respectively, to account for non-constant error precisions in the two sub-models.

4. Implementation using INLA

Bayesian hierarchical modelling that combines regression and error models has already been proposed by Lindley and El-Sayyad (1968), Clayton (1992) or Richardson and Gilks (1993). Recently, Muff et al. (2015) have shown how generalized linear (mixed) models can be fitted in the INLA framework when there is Berkson or classical ME in a Gaussian covariate. Here, we extend the methodology to the context of parametric survival modelling using the Weibull likelihood. At the same time, we allow for two-component error models with a mixture of Berkson and classical error in covariates, as well as for joint error modelling in multiple covariates. The hierarchical model encompasses the following levels:

(i) The Weibull *regression model* with hazard function (2), which defines the likelihood of the outcome $(\boldsymbol{t}, \boldsymbol{\delta})$ as a function of the error-free covariates $\boldsymbol{z}$, the true unobserved covariate(s) $\boldsymbol{x}$, and the shape parameter γ.

(ii) The *error model(s)*, either Berkson (4), classical (5), or a combination of both with levels (6) and (7), as described in the previous section. Error in multiple covariates can be modelled jointly, in which case the hierarchical model is augmented by all the respective equations.

(iii) A so-called *exposure model* (Gustafson, 2004), *i. e.* a model for each latent covariate $\boldsymbol{x}$, needs to be specified if the classical error model is used, or in the presence of Berkson *and* classical ME in $\boldsymbol{x}$. Here, we assume that the true covariate $\boldsymbol{x}$ is Gaussian with a mean that may depend on the error-free covariates in the matrix $\boldsymbol{z}$:

$$\boldsymbol{x} \,|\, \boldsymbol{z} \sim \mathrm{N}\left(\alpha_0 \boldsymbol{1} + \boldsymbol{z}\boldsymbol{\alpha}_z, \tau_x \boldsymbol{I}\right). \quad (8)$$

A special case emerges if $\boldsymbol{x}$ is independent of $\boldsymbol{z}$, *i. e.* when $\boldsymbol{\alpha}_z = \boldsymbol{0}$. Note that the above formulation assumes that the $\boldsymbol{x}$ are conditionally independent given

$\boldsymbol{z}$, although INLA is capable to deal with various Gaussian Markov random field (GMRF) dependencies within the latent field, such as temporal or spatial patterns. Still, an important prerequisite to apply INLA is the normality of $\boldsymbol{x} \,|\, \boldsymbol{z}$ (Rue et al., 2009), although recent extensions of INLA relaxed this restriction to near-Gaussian distributions (Martins and Rue, 2014). No exposure model is needed if a covariate suffers only from Berkson ME, because the error model (4) automatically specifies the distribution of $\boldsymbol{x} \,|\, \boldsymbol{w}$.

(iv) We assign independent Gaussian priors to the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_x, \boldsymbol{\beta}_z^\top)^\top$ and $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_z^\top)^\top$ and suitable gamma priors to the hyperparameters τ_{u_c}, τ_{u_b}, τ_x and γ. It is a particularity of INLA that β_x is also treated as a hyperparameter. The reason is that the term $x_i\beta_x$ in the regression equation is the product of two Gaussians, and is thus not Gaussian itself. For more details, see Muff et al. (2015) and additional explanations below.

For hierarchical models including a simple Berkson error, or a classical error with an exposure model that is independent of other covariates (*i. e.* $\boldsymbol{\alpha}_z = \mathbf{0}$ in (8)), the `meb` and the `mec` models (Muff et al., 2015) are available in `R-INLA`. However, all models discussed here feature a more complex structure, thus the explicit formulation of a joint model for the regression, error and exposure models is necessary. Suitably stacked predictor vectors and a corresponding response matrix are then required. Importantly, some of the equations in the joint model must be reformulated. For the novel classical/Berkson error model proposed here, equations (6) and (7) are written as

$$\begin{aligned} \mathbf{0} &= -\boldsymbol{r} + \boldsymbol{x} + \boldsymbol{u}_c \quad \text{and} \\ -\boldsymbol{w} &= -\boldsymbol{r} + \boldsymbol{u}_b \ , \end{aligned}$$

and the exposure model (8) for $\boldsymbol{x}$

$$\mathbf{0} = -\boldsymbol{x} + \alpha_0\mathbf{1} + \boldsymbol{z}\boldsymbol{\alpha}_z + \boldsymbol{\epsilon}_x \ ,$$

with suitably distributed vectors $\boldsymbol{u}_c$, $\boldsymbol{u}_b$ and $\boldsymbol{\epsilon}_x$. Thanks to this reformulation with pseudo-observations $\mathbf{0}$ or $-\boldsymbol{w}$ on the left, these equations can be interpreted as part of the observation model, which enables the joint fit of the hierarchical model in INLA. The response matrix in `R-INLA` then contains one separate column per equation,

namely

$$\begin{bmatrix} y_1 & \texttt{NA} & \texttt{NA} & \texttt{NA} \\ \vdots & \vdots & \vdots & \vdots \\ y_n & \texttt{NA} & \texttt{NA} & \texttt{NA} \\ \texttt{NA} & 0 & \texttt{NA} & \texttt{NA} \\ \vdots & \vdots & \vdots & \vdots \\ \texttt{NA} & 0 & \texttt{NA} & \texttt{NA} \\ \texttt{NA} & \texttt{NA} & -w_1 & \texttt{NA} \\ \vdots & \vdots & \vdots & \vdots \\ \texttt{NA} & \texttt{NA} & -w_n & \texttt{NA} \\ \texttt{NA} & \texttt{NA} & \texttt{NA} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \texttt{NA} & \texttt{NA} & \texttt{NA} & 0 \end{bmatrix},$$

where the first column corresponds to the regression model, columns two and three to the components of the error model, and the last column to the exposure model. The covariate vectors must be stacked in correspondence to the response matrix. Each equation of this joint observational model may follow a different likelihood function and requires a different set of hyperparameters.

An inherent complication of ME modelling in INLA is that product structures of two unknown parameters, such as $x_i\beta_x$ in (1), are generally not supported due to the requirement that the latent field must be (close to) Gaussian. As detailed out in Muff et al. (2015), a computational trick is then to interpret the regression parameter β_x as a hyperparameter, so that conditionally on β_x, the product $x_i\beta_x$ is still Gaussian. Moreover, the latent covariate $\boldsymbol{x}$ occurs not only as a product with β_x in the linear predictor, but also without a scaling in the error and exposure models. Identical copies of the same random field $\boldsymbol{x}$ are thus needed in different contexts, which can be implemented in INLA through the use of the copy-option (Martins et al., 2013). To facilitate the use of the error models presented here, we provide a more detailed description and `R-INLA`-code for a working example of the two-component error model in the Supplementary Material [Insert Inline Supplementary Computer Code "R-INLA_code_example_of_two_component_error_model.r" here].

5. Application: CVD mortality in the Swiss National Cohort Study

A major difference in the analysis carried out here to the original approach presented in Von Gunten et al. (2013) is the substitution of SBP by its transformation $\log(\text{SBP} - 50)$, and the replacement of the Cox by a Weibull regression model. For comparison, Cox and Weibull models were fitted with a (partial) maximum likelihood (ML) approach using the `coxph()` and `survreg()` functions from the `survival` pack-

age (Therneau, 2015) in `R` version 3.3.2 (R Core Team, 2016), and a Bayesian Weibull model using the `R` interface `R-INLA` (updated Jan 07, 2017) with the same prior distributions as will be specified later in Section 5.1.2. All continuous covariates were centered, but not standardized. The results (Table 3) indicate that the coefficients estimated for the Cox and the Weibull model are very similar for the NRP 1A data subset including the 3607 male participants used here, thus replacement of the Cox by the Weibull model seems warranted. In addition, a simultaneous p-value for the agreement of the cumulative baseline hazards between the Weibull (fitted with ML) and the Cox models was calculated from 10 000 bootstrap samples that were fitted with Cox regression, using the method described in Sabanés Bové and Held (2011, appendix C.1) and Held (2004). The resulting $p = 0.077$ yields only weak evidence for a conflict between Cox and Weibull regression. Figure 2 shows the 95% simultaneous credible bounds for the log cumulative baseline hazard function, as well as the respective curve from Weibull regression and a subset of 200 bootstrap samples fitted with Cox regression. The figure indicates that the Weibull curve stays within the credible bounds at all times. However, this analysis ignores the uncertainty in the estimated Weibull curve, which implies that the p-value reported above tends to be too optimistic, *i. e.* too small. Note that the output for the Bayesian Weibull model in Table 3 will serve as reference, hereafter called the *naive* estimates, to which the results after Bayesian error modelling will be compared.

As mentioned in Section 3.1, the advantage of Weibull regression is its interpretation as an AFT model, thus ETRs can be estimated as well. Estimates of the ETR values for the NRP 1A dataset obtained from the Weibull regression model fitted with INLA are given in Table 4 for continuous and binary covariates. For example, the ETR of 0.974 for blood glucose means that, as the blood glucose concentration increases by one mmol/L, the expected survival time of the patient decreases by 2.6%. To obtain credible intervals for the ETRs, which involve the ratio of two estimated parameters, the functions `inla.posterior.sample()` and `inla.hyperpar.sample()` generate Monte Carlo samples from the joint posterior distribution of the regression coefficients and the (hyper)parameters. For the covariate in the k-th column of $\boldsymbol{v}$, $k \geq 2$, Monte Carlo samples of the corresponding ETR can then easily be calculated as $\kappa_j^{(k)} = \exp(-\beta_j^{(k)}/\gamma_j)$, where $\beta_j^{(k)}$ and γ_j denote the respective sample values of iteration j. Posterior marginals, and in particular credible intervals, for the respective ETR can then be directly estimated from these samples. The estimates derived from INLA, using the same priors as above and no error modelling, are shown in Table 4. The table also contains the posterior probabilities $\mathsf{Pr}(\kappa \leq 1 \,|\, \text{data})$ that the ETR is below 1, the Bayesian analogue of a one-sided p-value. The main advantage of estimating the CIs for the ETRs using INLA in our application is that the same procedure can also be applied to hierarchical models including ME components.

	Cox		**Weibull (ML)**		**Weibull (Bayes)**	
	coef (se)	HR	coef (se)	HR	coef (sd)	HR
Intercept	-	-	-2.73 (0.15)	-	-2.64 (0.14)	-
Glucose	0.06 (0.03)	1.06	0.05 (0.03)	1.06	0.06 (0.03)	1.06
Age	0.13 (0.005)	1.13	0.12 (0.005)	1.13	0.12 (0.005)	1.12
Fasting time	-0.003 (0.01)	1.00	-0.002 (0.01)	1.00	-0.003 (0.01)	1.00
Phys. activity score	-0.05 (0.03)	0.95	-0.05 (0.03)	0.95	-0.05 (0.03)	0.95
BMI	0.02 (0.01)	1.02	0.02 (0.01)	1.02	0.02 (0.01)	1.02
Cholesterol	0.16 (0.07)	1.18	0.16 (0.07)	1.17	0.15 (0.07)	1.16
Mean $\log(\text{SBP} - 50)$	1.14 (0.24)	3.12	1.09 (0.24)	2.99	1.11 (0.24)	3.03
Cigarettes per day	0.027 (0.005)	1.03	0.027 (0.005)	1.03	0.025 (0.005)	1.03
3 main meals Yes	-0.11 (0.12)	0.90	-0.11 (0.12)	0.90	-0.13 (0.12)	0.87
Widowed	0.38 (0.21)	1.46	0.37 (0.21)	1.44	0.38 (0.21)	1.45
Divorced/separated	0.13 (0.24)	1.14	0.13 (0.24)	1.14	0.11 (0.24)	1.12
Single	0.26 (0.18)	1.29	0.24 (0.18)	1.27	0.21 (0.17)	1.23
Lugano	0.24 (0.19)	1.27	0.20 (0.19)	1.22	0.17 (0.19)	1.19
Nyon	-0.10 (0.12)	0.90	-0.10 (0.12)	0.90	-0.13 (0.11)	0.88
Solothurn	0.21 (0.16)	1.23	0.20 (0.16)	1.22	0.16 (0.16)	1.18
Vevey	-0.14 (0.19)	0.87	-0.15 (0.19)	0.86	-0.18 (0.18)	0.84
Shape parameter γ	-	-	2.04 (0.08)	-	2.02 (0.07)	-

Table 3: Comparison of the $\boldsymbol{\beta}$ estimates from the Cox and Weibull regression models fitted to the NRP 1A data. For the Weibull regression model, a maximum likelihood approach, as well as a Bayesian approach using INLA, were used for fitting. None of the calculations involved error modelling.

	κ	Equi-tailed 95% CI			$\mathsf{Pr}(\kappa \leq 1 \mid \text{data})$
Glucose	0.973	0.948	to	0.999	0.98
Age	0.944	0.939	to	0.948	1.00
Fasting time	0.836	0.677	to	1.020	0.96
Physical activity score	1.002	0.989	to	1.015	0.42
BMI	0.990	0.976	to	1.004	0.92
Cholesterol	0.928	0.869	to	0.990	0.99
Mean $\log(\text{SBP} - 50)$	0.582	0.459	to	0.729	1.00
No. of cigarettes per day	0.988	0.983	to	0.992	1.00
3 main meals	1.070	0.951	to	1.199	0.13

Table 4: Estimates of the ETRs (κ), equi-tailed 95% CIs and posterior probabilities $\mathsf{Pr}(\kappa \leq 1|\text{data})$ for all continuous or binary covariates calculated from 10 000 Monte Carlo samples of the INLA joint posteriors for the Weibull regression model without error modelling.

In the following subsections, we specify hierarchical error models for the reported mean number of cigarettes smoked per day and for the SBP, and report the results of fitting these models in `R-INLA`.

5.1. Error in the reported mean number of cigarettes smoked per day

The histogram of self-reported mean number of cigarettes smoked per day in the NRP 1A study is depicted in the left panel of Figure 1. Wang et al. (2012), who related

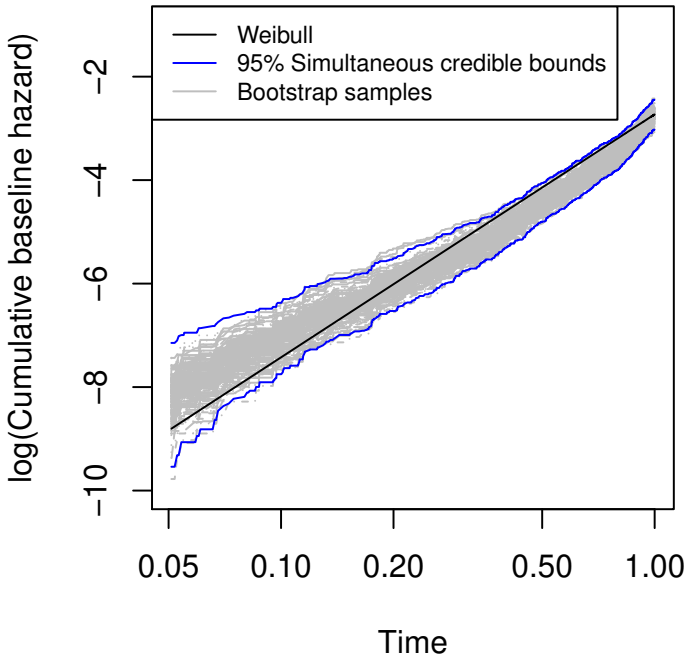

Figure 2: Log cumulative baseline hazard $\beta_0 + \gamma \log(t)$ for the Weibull regression model, using the estimates for β_0 and γ from the ML output (black line), as well as 95% simultaneous credible bounds for the cumulative baseline hazard of the Cox model (blue lines). Grey lines show 200 out of 10 000 bootstrap samples using Cox regression, given here for illustration.

true cigarette counts that were assessed by ecological momentary assessment (EMA), *i. e.* the instantaneous recording of each cigarette as it was smoked, to retrospective records obtained by time-line follow-back (TLFB) methods, formulated an error model with two levels to account for both the rounding and the misremembering component in the reported cigarette numbers.

5.1.1. Error model

A two-component error model accounting for misremembering (classical error) and rounding (Berkson error) in the cigarette reports of smoking individuals was therefore specified, using equations (6) and (7). Note that this corresponds to a different modelling strategy in comparison to Wang et al. (2012), where daily cigarette counts were considered, while we are looking at average daily consumption, which may involve floating point values. Moreover, our model does not directly allow for biased reports, such as over- or underreporting of the true numbers, although such extensions could be handled by adding respective terms to the models. In addition to the two error modelling layers, an exposure model for the latent true smoking covariate $\boldsymbol{x}$ is needed.

Assuming that the smoking behavior is not influenced by any other covariates, a normally distributed, independent exposure model with an average value μ_x and precision τ_x was used for the smoking individuals. Note that the non-smokers must be treated separately, as we assumed here that they reported their consumption correctly (*i. e.*, $w_i = 0$ implies $x_i = 0$). We thus have

$$x_i \sim \mathrm{N}(\mu_x, \tau_x) \ , \tag{9}$$

if individual i was a smoker, *i. e.* reported $w_i > 0$.

5.1.2. Priors

The specification of informative priors is important in the context of error modelling, because the ensemble of all parameters is generally nonidentifiable (Gustafson, 2005). We therefore need to carefully evaluate prior information from external validation data and expert knowledge.

misremembering component (Classical error). We start with the determination of the prior for τ_{u_c} and the entries $d_i^{(c)}$ in the scaling matrix as given in equation (6). Heteroscedastic error modelling is important because it has been observed previously that lighter smokers reported their cigarette consumption with higher precision (Klesges et al., 1995). It is thus crucial to allow the scaling constants to depend on the reported values w_i. To obtain the respective information, we used the results presented by Wang et al. (2012), where true daily smoked values x_i from EMA assessments were compared to retrospectively reported values w_i. Wang et al. (2012) assumed that the remembered values r_i (before rounding) are Poisson distributed as

$$\mathsf{E}(r_i \,|\, x_i, b_i) = \exp(\beta_0 + \log(x_i)\beta_1 + b_i) \ ,$$

with a random effect $b_i \sim \mathcal{N}(0, \tau_b)$, and estimated the parameters as $\beta_0 = 2.32$, $\beta_1 = 0.27$ and $\tau_b = 1/0.09$. As the true values x_i and the remembered values r_i are not assumed to be integer numbers here, we only used this error model to obtain prior information for the error precision τ_{u_c}. To this end, 10 000 remembered values $\tilde{\boldsymbol{r}}_x$ for any true integer value x between 1 and 80 were simulated. Figure 3 shows that the error variance increases linearly with the square root of smoking counts, and thus the precision was assumed to scale as $d_i^{(c)} \propto \sqrt{1/w_i}$ when individual i reported $w_i > 0$. For $x = 20$, for example, the error variance was estimated to be $\mathrm{Var}(\tilde{\boldsymbol{r}}_{20} - 20 \cdot \mathbf{1}) = 79$. By setting the scaling factor $d_i^{(c)} = 1$ for reported values $w_i = 20$, we can use the $\tau_{u_c} \sim \mathrm{G}(1/79, 1)$ prior for the error precision, with mean equal to $1/79$, and a variance that is identical to the mean. Consequently, the scaling for general $w_i > 0$ is $d_i^{(c)} = \sqrt{20/w_i}$. Finally, non-smokers were assumed to report correct values, thus $d_i^{(c)} = 10^{15}$ was used for $w_i = 0$ to avoid numerical problems.

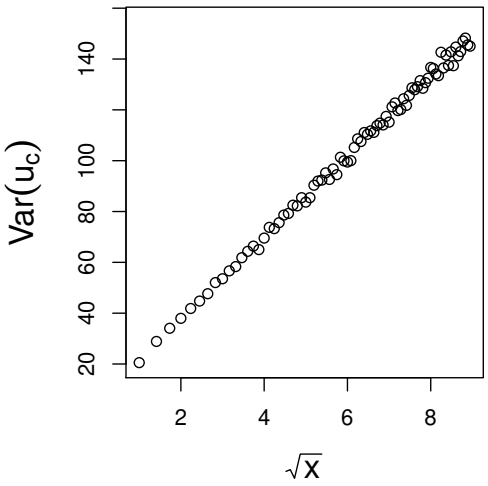

Figure 3: The misremembering error variance $\mathrm{Var}(\boldsymbol{u}_c)$ for the error model of self-reported cigarette consumption increases linearly with the square root of the true smoking value. The variances were estimated from simulations using the model of Wang et al. (2012).

Rounding component (Berkson error). To determine the priors for the error precision τ_{u_b} and the scaling factors $d_i^{(b)}$ of model (7), we start with the prior distribution of the error when the reported number was $w_i = 10$. For a lower bound of the precision we assumed that the remembered values lie within a range of plus or minus 5, meaning that they are somewhere in between 5 and 15, as they would otherwise very likely be rounded to 5, 15 or 20. Assuming a uniform ("worst-case") distribution for the remembered values, the lower bound for the precision is then given by 12/100. On the other hand, the minimal spread around the reported value is assumed to be plus or minus 1, leading to an upper bound of the precision being 3. Using these values as 2.5% and 97.5% quantiles of a gamma distribution, numerical optimization leads to the $\tau_{u_b} \sim \mathrm{G}(1.92, 1.81)$ prior in the case of $w_i = 10$. To determine the scaling factors $d_i^{(b)}$, we assumed the same scaling behavior as for the misremembering error, *i. e.* that the error precision increases with $\sqrt{1/w_i}$, if the value is a multiple of five. By setting the scaling factor to $d_i^{(b)} = 1$ for $w_i = 10$, we obtain the formula $d_i^{(b)} = \sqrt{10/w_i}$ for multiples of five. For all other values we suppose that CIs are only half as wide, resulting in a four times larger precision $d_i^{(b)} = 4 \cdot \sqrt{10/w_i}$. Again, reported values by nonsmokers, *i. e.* $w_i = 0$, were assumed to be error-free, thus $d_i^{(b)} = 10^{15}$ was used. In summary, the scaling is given as

$$d_i^{(b)} = \begin{cases} 10^{15} & \text{if } w_i = 0 \ , \\ \sqrt{10/w_i} & \text{if } w_i \bmod 5 \equiv 0 \ , \\ 4 \cdot \sqrt{10/w_i} & \text{otherwise} \ . \end{cases}$$

Other priors. In the exposure model (9), the mean was fixed at $\mu_x = 17.98$, which corresponds to the average reported number by the smokers in the study, and the

prior for the precision of the exposure model (9) was set to $\tau_x \sim \mathrm{G}(0.0097, 1)$, with expected value and variance equal to the sampling variance calculated from the 3176 smoking individuals of the EMA records in Wang et al. (2012). For the components of $\boldsymbol{\beta}$ in the regression model (1), we assumed independent normal priors inspired by the g-prior (Zellner, 1986) for the regression coefficients in the normal linear model. Let $\boldsymbol{v}_c$ denote the design matrix of the transformed and centered observed covariates including the error-prone covariates, with ones in the first column representing the intercept. Moreover, let n_{obs} be the number of observed deaths due to CVD, here $n_{\text{obs}} = 458$. It was then assumed that the prior variance of the k-th covariate is

$$\mathrm{Var}(\beta^{(k)}) = n_{\text{obs}} \left[(\boldsymbol{v}_c^\top \boldsymbol{v}_c)^{-1} \right]_{kk},$$

where $[.]_{kk}$ is the k-th diagonal entry of the matrix. The choice of n_{obs} (rather than n) is motivated by extensions of BIC to survival models (Volinsky and Raftery, 2000), although the results are not sensitive to this choice, as illustrated by a sensitivity check in the Supplementary Material (Section 3). Assuming prior means of zero, the prior distributions thus were

$$\beta^{(k)} \sim \mathrm{N}\left(0, \left\{ n_{\text{obs}} \left[(\boldsymbol{v}_c^\top \boldsymbol{v}_c)^{-1} \right]_{kk} \right\}^{-1}\right) . \tag{10}$$

Alternatively, a multivariate prior

$$\boldsymbol{\beta} \sim \mathrm{N}\left(\mathbf{0}, n_{\text{obs}}^{-1}\ \boldsymbol{v}_c^\top \boldsymbol{v}_c\right) \tag{11}$$

could have been used. It is not straightforward, but possible to implement such a prior with a fixed covariance matrix in `R-INLA` (Held and Sauter, 2016). However, our independent prior formulation (10) shares with the full g-prior (11) the attractive property of measurement invariance under rescaling and translation, so all analyses done here were based on independent priors. Another alternative, inspired by generalized g-priors for GLMs (Wang and George, 2007; Li and Clyde, 2016), uses the observed Fisher information matrix from the Weibull regression model instead of $\boldsymbol{v}_c^\top \boldsymbol{v}_c$, which leads to similar prior precisions as in (10). Note however that the prior based on the observed Fisher information is outcome-dependent. For the shape parameter γ of the Weibull distribution, we used a gamma prior with mean 2, which is similar to what was found in the naive regression analysis ($\hat{\gamma} = 2.02$, see Table 3) and corresponds to a linearly increasing expected baseline hazard function. Our prior thus had the form $\mathrm{G}(2a, a)$ for some $a > 0$. In addition, assuming that a decreasing hazard is implausible, given the age cohort and the follow-up time of more than 30 years, the probability for a decreasing hazard was set to $\Pr(\gamma < 1) = 0.025$, which led to $a \approx 5.5$ and thus to the $\gamma \sim \mathrm{G}(11, 5.5)$ prior.

5.1.3. Results

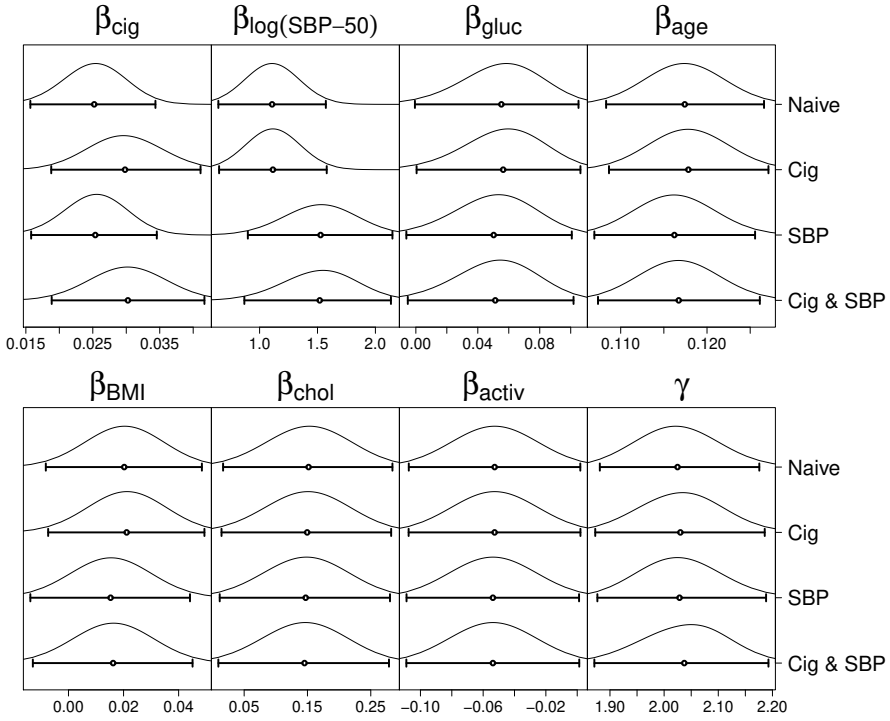

Figure 4: Posterior marginal distributions (thin lines) and 95% CIs with posterior mean estimates (point on the CI) for all regression coefficients whose respective covariate was either subject to error modelling, or that was included as a covariate in the exposure model of SBP, as well as the shape parameter γ. The four models were the naive without any error considerations (Naive), a two-component classical/Berkson error to account for the error in the reported cigarette consumption (Cig), a two-component classical/Berkson error model for the blood pressure (SBP), and a joint model for the error in both covariates (Cig & SBP).

The estimates and 95% CI for some parameters of interest are labelled as Cig in Figure 4. The estimate of the regression coefficient β_{cig} for the mean number of cigarettes smoked per day slightly increased from $\hat{\beta}_{\text{cig}} = 0.025$ (with 95% CI from 0.016 to 0.034) in the naive model to $\hat{\beta}_{\text{cig}} = 0.030$ (with 95% CI from 0.019 to 0.041) after error correction. The shape parameter γ was again estimated as 2.03 (95% CI from 1.87 to 2.19). The classical error precision for misremembering of the mean number of cigarettes smoked per day was $\hat{\tau}_{u_c} = 0.025$ (with 95% CI from 0.018 to 0.036), and the rounding error precision $\hat{\tau}_{u_b} = 1.16$ (with 95% CI from 0.23 to 3.12). Posterior means and CIs of the most relevant ETRs from 10 000 MCMC samples are graphically displayed in Figure 5. In particular, the ETR for a single cigarette decreased from 0.988 to 0.985. This implies that for an average daily consumption of 20 cigarettes, the expected life time shrinks by a factor 0.75, compared to 0.78 without error modelling. However and importantly, the coefficient and ETR for plasma glucose concentration - the main explanatory variable in the study by Von Gunten et al. (2013) - were essentially unaffected by modelling the ME in the reported cigarette values.

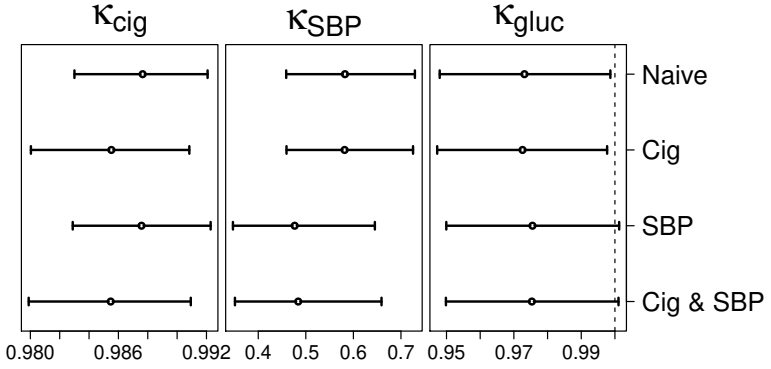

Figure 5: Estimates and 95% CIs for the ETRs of the mean number of cigarettes smoked per day (κ_{cig}), for the mean transformed systolic blood pressure (κ_{SBP}), and for blood glucose concentration (κ_{gluc}). The estimates were derived by Monte Carlo sampling from the posterior distributions given by INLA.

5.2. Error in systolic blood pressure measurements

Due to daily variations and imprecision in the measurement devices, the long-term SBP cannot be precisely quantified. Single measurements at health examinations have to be used as proxies for the long-term SBP, which may differ considerably (Carroll et al., 2006). In addition, end-digit preference of SBP reportings are blurring the daily measurements.

5.2.1. Error model

As in the case of cigarette misreporting, a two-component error model was formulated to account for measurement imprecision (classical error) and rounding/digit preference (Berkson error) in SBP reportings. We denote the j-th SBP measurement for participant i by w_{ij} and the vector of all j-th measurements is given by $\boldsymbol{w}_j$, with the i-th component encoded as missing value (`NA`), if the measurement w_{ij} was not available.

Let $\overline{\boldsymbol{w}}$ be the vector of the means over all available (transformed) SBP measurements per participant, which corresponds to the covariate included in the original study without error modelling. The two-component error model for the observed vector $\overline{\boldsymbol{w}}$ is then given as

$$\overline{\boldsymbol{r}} = \boldsymbol{x} + \overline{\boldsymbol{u}}_c, \quad \overline{\boldsymbol{u}}_c \sim \mathrm{N}(0, \tau_{u_c}\boldsymbol{D}_c) \quad \text{and} \tag{12}$$

$$\overline{\boldsymbol{r}} = \overline{\boldsymbol{w}} + \overline{\boldsymbol{u}}_b, \quad \overline{\boldsymbol{u}}_b \sim \mathrm{N}(0, \tau_{u_b}\boldsymbol{D}_b) \ , \tag{13}$$

where $\overline{\boldsymbol{u}}_c$ and $\overline{\boldsymbol{u}}_b$ denote the vector of mean classical and Berkson error terms of the individual measurements for each participant, and $\overline{\boldsymbol{r}}$ represents the mean of the values indicated by the measurement device *before* they were rounded. In addition, a model for the latent SBP covariate is needed. As in Muff et al. (2015) we used an exposure

model that allowed for linear dependency on other covariates. Based on subject matter knowledge, the following six covariates were selected: blood glucose level, age at study entry, BMI, blood cholesterol level, physical activity score and reported mean number of cigarettes smoked per day. The exposure model can then be written as

$$\boldsymbol{x} = \alpha_0 \mathbf{1} + \widetilde{\boldsymbol{z}}\boldsymbol{\alpha}_{\widetilde{z}} + \boldsymbol{\epsilon}_x, \quad \boldsymbol{\epsilon}_x \sim \mathrm{N}(\mathbf{0}, \tau_x \boldsymbol{I}) \ , \tag{14}$$

where $\widetilde{\boldsymbol{z}}$ is the submatrix containing the columns corresponding to the six covariates mentioned above, and $\boldsymbol{\alpha}_{\widetilde{z}} = (\alpha_1, \ldots, \alpha_6)^\top$ is the vector of the respective regression coefficients.

5.2.2. Priors

Measurement imprecision component (Classical error). The error precision prior for a single measurement was set to $\tau_{u_c} \sim \mathrm{G}(100, 1)$, in analogy to Muff et al. (2015, application 5.2), where the same prior has been derived for the classical error component from expert knowledge. As averaging over more measurements reduces the variance of the mean error term, the diagonal entries $d_i^{(c)}$ in $\boldsymbol{D}_c$ had to be scaled based on the number of measurements per individual. Under the assumption that the included measurements 1, 3 and 4 suffer from independent ME and using that thus the error precision increases linearly with the number of repeats n_i for patient i, the scaling constant was given as $d_i^{(c)} = n_i$.

Rounding component (Berkson error). Similar considerations as for the rounding phenomenon in cigarette numbers are useful when determining priors for the rounding error caused by digit preference in SBP reports. Here, an additional complication comes from the fact that $\log(\mathrm{SBP} - 50)$ instead of SBP were used in the regression model. Let us start by considering the case when the reported value was SBP=130, which is close to the mean of the observations in the present dataset. For a lower bound of the precision we assume that the actual measurements lie within a range of plus or minus 10, that is, that the measurement device showed a value between 120 and 140. On the transformed scale, we thus assume that all values between $\log(130-50-10)$ and $\log(130-50+10)$ are equally likely, implying a precision of 190. At the other extreme it was assumed that the actual value lies between 129 and 131, with uniform distribution between $\log(130-50-1)$ and $\log(130-50+1)$ on the transformed scale, which leads to a maximal precision of 19 198. Using these two extremes as the 2.5% and 97.5% quantiles of a gamma distribution, numerical optimization leads to $\tau_{u_b} \sim \mathrm{G}(1.113, 0.00020)$ for the case of SBP_i=130.

In the second step the scaling factors $d_i^{(b)}$ need to be determined. Note that the effect of adding or subtracting a constant depends on the actual SBP values due to the log transformation. Starting our considerations with observed SBP values with end-

digts zero $(100, 110, 120, \ldots)$, and assuming the same rounding behavior as for 130, the precision was scaled with $1/(\log(\text{SBP}_i - 50 + 10) - \log(\text{SBP}_i - 50 - 10))^2$, the inverse of the squared width of the respective interval. Next, because end-digits 2, 4, 5, 6 and 8 had a similar frequency (Figure 1, right), they were scaled with the same constant, assuming that the actual measurement spread no more than within the range of plus or minus 2 units in these cases. Finally, for the remaining end-digits (1, 3, 7 and 9) a maximum spread of plus or minus 1 unit was assumed. In summary, the scaling was thus given as

$$d_i^{(b)} = \begin{cases} [c/(\log(\text{SBP}_i - 50 + 10) - \log(\text{SBP}_i - 50 - 10))]^2 & \text{for end-digit 0 ,} \\ [c/(\log(\text{SBP}_i - 50 + 2) - \log(\text{SBP}_i - 50 - 2))]^2 & \text{for end-digits } 2, 4, 5, 6, 8 \text{ ,} \\ [c/(\log(\text{SBP}_i - 50 + 1) - \log(\text{SBP}_i - 50 - 1))]^2 & \text{for end-digits } 1, 3, 7, 9 \text{ ,} \end{cases}$$

where $c = \log(130-50+10) - \log(130-50-10)$ was used to obtain $d_i^{(b)} = 1$ for the case SBP_i=130. In addition, the scaling accounted for cases where multiple measurements of SBP were available ($n_i > 1$), such that $d_i^{(b)} = n_i/(\sum_j 1/d_{ij}^{(b)})$, where $d_{ij}^{(b)}$ is the respective scaling of the j^{th} measurement of the i^{th} patient.

Other priors. The hyperprior for τ_x was set to $\tau_x \sim \text{G}(10, 1)$, as described in Muff et al. (2015, application 5.2). Priors for the components of $\boldsymbol{\beta}$ were again determined as in (10). Finally, for the parameters $\boldsymbol{\alpha} = (\alpha_0, \boldsymbol{\alpha}_{\tilde{z}}^\top)^\top$ of the exposure model we used a similar strategy as for the coefficients in the regression model: We chose Gaussian priors with mean zero inspired by the g-prior in the normal linear model. In the g-prior, the covariance matrix is proportional to $n \cdot \tau_x^{-1}$. Here, the variance τ_x^{-1} is unknown and we replace it by its prior expectation $\mathsf{E}[\tau_x^{-1}] = 1/9$ to obtain a prior that does not depend on other hyperparameters. Given the reduced design matrix $\widetilde{\boldsymbol{v}}_c = [\mathbf{1}, \widetilde{\boldsymbol{z}}]$ consisting of the intercept and the covariates included in the exposure model only, the prior distributions were

$$\alpha_k \sim \text{N}\left(0, \left\{\frac{n}{9}\left[(\widetilde{\boldsymbol{v}}_c{}^\top \widetilde{\boldsymbol{v}}_c)^{-1}\right]_{kk}\right\}^{-1}\right), \quad k = 0, \ldots, 6 \text{ ,}$$

assuming prior independence of α_k, $k = 1, \ldots, 6$.

5.2.3. Results

Estimates of the regression coefficients and the hyperparameters, together with the corresponding 95% CIs, are shown in Figure 4. The estimate of the regression coefficient β_{SBP} increased from the naive estimate $\hat{\beta}_{\text{SBP}} = 1.11$ (95% CI from 0.64 to 1.57), to $\hat{\beta}_{\text{SBP}} = 1.53$ (95% CI from 0.9 to 2.15). This corresponds to a change in the expected hazard ratio $\exp(\hat{\beta}_{\text{SBP}})$ from 3.03 to 4.60. Note, however, that also the width of the CIs increased, *i. e.* error modelling added more uncertainty to the estimates. The shape

parameter was estimated to be $\hat{\gamma}$=2.03 (95% CI from 1.88 to 2.19). Some estimates of ETRs after error-adjustment of SBP are displayed in Figure 5. The ETR for the blood pressure decreased from 0.58 (95% CI from 0.46 to 0.73) in the naive model to 0.48 (95% CI from 0.35 to 0.65). The interpretation of the ETR for the transformed SBP values is facilitated when specific values are plugged into the formula. For instance, while an increase in blood pressure from 120 to 160 mm Hg leads to an expected reduction in survival time by a factor 0.78 when using the estimates of the naive model, the error-corrected model implies a factor of 0.71. On the other hand, as when modelling the error in the reported cigarette consumption, the slope coefficient and ETR for plasma glucose concentration were essentially unaffected by modelling the ME in SBP.

Additional regression coefficients $\boldsymbol{\alpha}$ from the exposure model were estimated, and these results are given in Table 5. We conclude that all variables in the exposure model except the number of cigarettes smoked per day seem to be associated with the SBP. Note that including the covariates age and BMI in the exposure model for the SBP also had an impact on the estimates of the coefficients β_{age} and β_{BMI} in the regression model (see Figure 4). The estimated influence of these covariates was slightly corrected downward, probably in compensation to the increased influence that is now attributed to SBP.

	Coefficient	Equi-tailed 95% CI		
Intercept	0.004	-0.002	to	0.011
Glucose	0.014	0.009	to	0.019
Age	0.004	0.003	to	0.004
BMI	0.012	0.010	to	0.014
Cholesterol	0.013	0.003	to	0.023
Physical activity score	0.005	0.0009	to	0.009
No. of cigarettes per day	-0.00008	-0.0007	to	0.0005

Table 5: Estimated coefficients $\boldsymbol{\alpha}_{\tilde{z}}$ in the exposure model for SBP.

5.3. Joint error modelling in both covariates

Combining error models for several covariates in a Bayesian hierarchical model is conceptually straightforward, and once the models for error in single covariates have been formulated, the respective implementation is simple thanks to the modular framework of INLA. Here, we have combined the two-component error model for the mean number of cigarettes smoked per day described in Section 5.1 with the classical error model for the SBP of Section 5.2. However, the number of cigarettes smoked per day was excluded from the exposure model, which seemed not to be associated with the SBP (Table 5), to avoid the complication of an error prone covariate in the exposure model of another covariate.

The resulting model had seven levels, namely the regression model (2), the two components (6) and (7) for the cigarette misremembering and rounding error, the exposure model for the true cigarette numbers (9), the two-component error model (12) and (13) for the SBP, and the exposure model for the true SBP (14). All priors were identical to those derived in Sections 5.1 and 5.2.

5.3.1. Results

Joint error modelling led to very similar results compared to the models that were formulated for single covariates. Namely the effects related to cigarette consumption $\hat{\beta}_{\text{cig}} = 0.030$ (with 95% CI from 0.019 to 0.042) and SBP $\hat{\beta}_{\text{SBP}} = 1.52$ (with 95% CI from 0.87 to 2.13), as well as the shape parameter $\hat{\gamma}$=2.04 (95% CI from 1.87 to 2.19) were essentially identical, although with larger uncertainty than when the errors were modelled individually. Consequently, also the ETRs were similar among the models (Figure 5). The results from simultaneously modelling the error in SBP and in the mean cigarette consumption are thus consistent to those from the single covariate error models.

5.4. Sensitivity Analysis

As mentioned in Section 5.1.2, the parameters of error models are often nonidentifiable, thus the results from the error modelling procedure are expected to be sensitive to changes in the prior distribution. In the classical error component of the model given by equation (6), for instance, the variances of the classical error $\boldsymbol{u}_c$ and of the covariate $\boldsymbol{x}$ are confounded. On the other hand, Berkson error in covariates does generally have less severe effects on the estimates of regression coefficients, especially in the case of log-linear regression models (Carroll et al., 2006), thus the results are expected to be less sensitive on the prior for τ_{u_b}. For illustration a sensitivity check was carried out for the model incorporating error in SBP from Section 5.2. All results are presented in the Supplementary Material, Section 3. As expected, while the results change with variations in the prior on τ_{u_c}, there is no or only very low sensitivity on the prior choice for the regression parameters ($\boldsymbol{\beta}$, $\boldsymbol{\alpha}$) and for the Berkson error precision τ_{u_b}.

5.5. Comparison to MCMC results

In previous applications, INLA has frequently been compared to results from MCMC sampling procedures, such as OpenBugs (Lunn et al., 2009) or JAGS (Plummer, 2003), and it has been shown that ME modelling with INLA leads to consistent results (Muff et al., 2015). Therefore, only a brief comparison between the INLA results and a sample of 100 000 iterations (with a burn-in of 10 000) from an MCMC procedure generated via the `R`-interface `rjags` (Plummer, 2003; R Core Team, 2016) is given here, namely for the Berkson/classical error model in SBP from Section 5.2. The INLA and MCMC posterior distributions for $\hat{\beta}_{\text{SBP}}$ are in almost perfect agreement, as shown in Figure

6 and Section 4 of the Supplementary Material, and the same quality of the INLA approximation is expected for the other models. In terms of efficiency, the respective model could be fitted in roughly 4 minutes on an IntelCore i7-2640M 2.80GHz processor, while MCMC sampling required more than 9 hours.

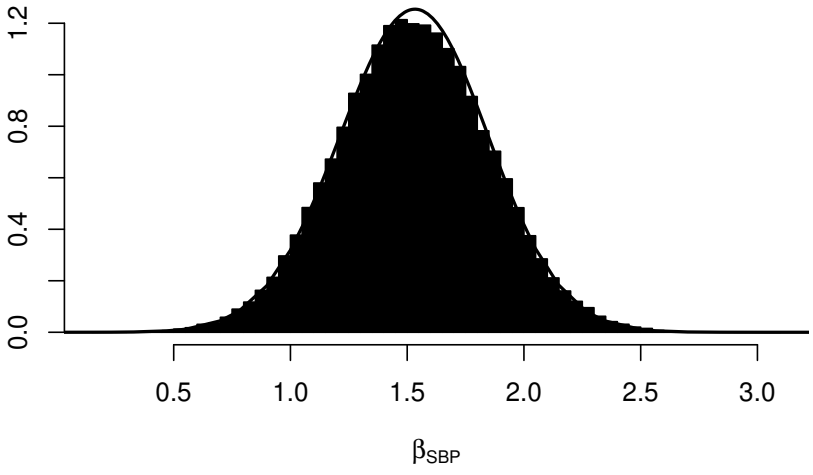

Figure 6: Comparison of MCMC samples (histogram) with the posterior marginal for $\hat{\beta}_{SBP}$ from INLA for the Berkson/classical error model in SBP.

6. Discussion

The INLA framework has previously been shown to be a very powerful tool for Bayesian hierarchical modelling, in particular for GLMMs (Fong et al., 2010), but also in the context of survival regression (Martino et al., 2011). Recently, INLA has been used to fit simple error models for continuous covariates in a GLMM context (Muff et al., 2015). This promising methodology has been extended here in three ways: First, we have accounted for a multi-component error structure, namely when superimposed Berkson and classical error components obscure a true covariate simultaneously. Second, we have jointly modelled ME in multiple covariates. And third, we have shown that Bayesian ME modelling in Weibull regression models is also straightforward.

Using this improved methodology we have reanalyzed the subset of male participants from the NRP 1A study, which investigated the factors that lead to increased CVD mortality (Von Gunten et al., 2013). Accounting for a two-component (classical and Berkson) ME in self-reported average cigarette consumption or in SBP, respectively, increased the corresponding estimated regression coefficients. Moreover, error modelling also affected slope estimates of other covariates. In turn, this means that

the inclusion of the error-prone covariates may lead to biased parameter estimates, an aspect that has already been discussed for Cox regression in the case of classical covariate error (Augustin and Schwarz, 2002; Augustin, 2004). Moreover, attenuation effects induced by classical error in covariates are known from various other types of regressions models. Importantly, also reverse attenuation (*i. e.* overestimation) of effects may occur, in particular when error-prone covariates are correlated with other predictors (Fuller, 1987; Carroll et al., 2006). Here, ME modelling in SBP led to a slight downward correction for the estimated effects of age and BMI (Figure 4), which are indeed related to the blood pressure according to the results in Table 5. Modelling the ME in the cigarette counts also slightly decreased the estimated effect of age. On the other hand, the presence of additive Berkson ME in a covariate of a Cox regression model was found to have only negligible effects on parameter estimates, provided that the error variance is not too large (Küchenhoff et al., 2007). In fact, when the errors in the self-reported cigarette counts were only treated as a rounding and thus as a Berkson error, while neglecting the misremembering (*i. e.* classical) component, we found essentially no changes in the estimated parameters in comparison to the naive model (results not shown). This also illustrates that the results are sensitive to the choice of the error model, and that a good understanding of the error structure is important. Here, we borrowed prior knowledge from external validation data for cigarette reporting behavior (Wang et al., 2012) and digit preference in SBP (de Lusignan et al., 2004), as well as from previously derived priors using expert knowledge (Muff et al., 2015). Notably, all results are qualitatively consistent when the subset of female participants was analyzed (results not shown). We conclude that especially classical error components, as those present in the blood pressure measurements or in the misremembering component of cigarette self-reporting, should not be ignored. Given that both covariates are frequently included as confounders, or even as main predictor in medical studies, and that good prior knowledge of the error distributions exists, we suggest that error modelling for these covariates should generally be considered.

Note that one concern for SBP measurements is the occurrence of the so-called *white coat hypertension effect* (Khan et al., 2007), which implies that the first measurement during a consultation often overestimates the actual patient's SBP. Given that in this study either only one measurement was available per consultation, or that the second measurement had to be discarded (*i. e.*, in the first consultation), the recorded mean SBP values might overestimate the actual long-term SBP. To account for such a systematic effect, the classical error model component given in equation (12) can by extended by adding a constant vector $\boldsymbol{\delta}$ such that $\overline{\boldsymbol{r}} = \boldsymbol{\delta} + \boldsymbol{x} + \overline{\boldsymbol{u}}_c$, where the entries of $\boldsymbol{\delta}$ represent the expected overestimation. Such an extension is however not considered here, because it would merely relocate (shift) the variable, which affects the estimate of intercept (β_0), but not the slope parameters of interest. Of course, things become

more complicated if the entries of $\boldsymbol{\delta}$ depend on $\boldsymbol{x}$ or any other variables, such as sex or age, but these are modelling aspects that are not in the scope of the present paper.

Bayesian approaches for ME modelling in survival models have so far not been very popular, an interesting exception being Tadesse et al. (2005), who proposed MCMC sampling to account for classical covariate error in Cox regression. One advantage of the INLA approach presented here is that no sampling is required to obtain parameter estimates, and that models can thus be fitted efficiently and with high accuracy. Interesting extensions of the error models presented here can be handled as well. For instance, Muff and Keller (2015) not only included heteroscedastic error variances, but also heteroscedastic variances for the unobserved covariate $\boldsymbol{x}$. In addition, the Gaussian assumption for the classical error term may by replaced by, *e. g.* a Poisson or negative binomial model. Such extensions are possible, as long as the latent covariate $\boldsymbol{x}$ still has a GMRF structure, which, among others, may involve spatial or temporal dependencies. Recent extensions of INLA relaxed the Gaussian restriction for the latent field to near-Gaussian distributions, such as the Student t distribution (Martins and Rue, 2014). Such extensions may also be useful when random effects or *frailties* (*e. g.* subject-specific covariate effects and intercept) are added to the linear predictor of survival models, in order to obtain robust inference in the presence of outliers, as has been shown by McCrink (2016).

Note that there are certain limitations to the model complexity that INLA is able to cope with. Although arbitrarily complex models can be formulated in principle, there are practical limitations to the number of hyperparameters, as numerical integration needs to be performed over the space of all hyperparameters (Rue et al., 2009). Here, the model fitting procedure for the error models involving only the smoking covariate or only the SBP, respectively, required 5 hyperparameters. On the other hand, the joint error model required 9 hyperparameters, and the calculation took approximately one hour, thus significantly longer than for the two models that only accounted for error in one covariate at the time (4 minutes each). Moreover, our example illustrated that fitting a joint model does not necessarily provide much additional insight with respect to sequential error modelling of single covariates, although this might be different in other applications. Still, if a joint model fit is too demanding, combining the results from error modelling in single covariates can lead to useful approximations, in particular if the error-prone covariates are not correlated.

Acknowledgements

We thank Prof. Saul Shiffman for the permission to use their data set on true vs. reported cigarette consumption. This work would not have been possible without generous funding from the Faculty of Science of the University of Zurich.

Augustin, T., 2004. An exact corrected log-likelihood function for cox's proportional hazards model under measurement error and some extensions. Scandinavian Journal of Statistics 31, 43–50.

Augustin, T., Schwarz, R., 2002. Cox's proportional hazards model under covariate measurement error: A review and comparison of methods. In: Van Huffel, S., P., L. (Eds.), Total Least Squares and Error-in-Variables Modelling: Analysis, Algorithms and Applications. Springer Netherlands, pp. 175–188.

Berkson, J., 1950. Are there two regressions? Journal of the American Statistical Association 45, 164–180.

Bopp, M., Braun, J., Gutzwiller, F., Faeh, D., for the Swiss National Cohort Study Group, 2012. Health risk or resource? gradual and independent association between self-rated health and mortality persists over 30 years. PLoS ONE 7, e30795, DOI: 10.1371/journal.pone.0030795.

Camarda, C. G., Eilers, P. H. C., Gampe, J., 2008. Modelling general patterns of digit preference. Statistical Modelling 8, 385–401.

Carroll, K. J., 2003. On the use and utility of the Weibull model in the analysis of survival data. Controlled Clinical Trials 24, 682–701.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective, 2nd Edition. Chapman & Hall, Boca Raton.

Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T., Abbott, R. D., 1984. On errors-in-variables for binary regression models. Biometrika 71, 19–25.

Carroll, R. J., Stefanski, L. A., 1990. Approximate quasilikelihood estimation in models with surrogate predictors. Journal of the American Statistical Association 85, 652–663.

Clayton, D., 1992. Models for the longitudinal analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer, J. H., Feinleib, M., Lippert, P., Hoffmeister, H. (Eds.), Statistical Models for Longitudinal Studies of Health. Oxford University Press, pp. 301–331.

Collett, D., 2015. Modelling Survival Data in Medical Research, 3rd Edition. Texts in Statistical Science. Chapman & Hall/CRC, Boca Raton.

Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association 89, 1314–1328.

Cox, D. R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 34, 187–220.

de Lusignan, S., Belsey, J., Hague, N., Dzregah, B., 2004. End-digit preference in blood pressure recordings of patients with ischaemic heart disease in primary care. Journal of Human Hypertension 18, 261–265.

Dellaportas, P., Stephens, D. A., 1995. Analysis of errors-in-variables regression models. Biometrics 51, 1085–1095.

Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian inference for generalized linear mixed models. Biostatistics 11, 397–412.

Fuller, W. A., 1987. Measurement Error Models. John Wiley & Sons, New York.

Gimenez, P., Bolfarine, H., Colosimo, E. A., 1999. Estimation in Weibull regression model with measurement error. Communications in Statistics, Theory & Methods 28, 495–510.

Gleser, L. J., 1990. Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In: Brown, P. J., Fuller, W. A. (Eds.), Statistical Analysis of Measurement Error Models and Applications. American Mathematical Society, pp. 99–114.

Gustafson, P., 2004. Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. Chapman & Hall/CRC, Boca Raton.

Gustafson, P., 2005. On Model Expansion, Model Contraction, Identifiability and Prior Information: Two Illustrative Scenarios Involving Mismeasured Variables. Statistical Science 20, 111–140.

Gutzwiller, F., Nater, B., Martin, J., 1985. Community-based primary prevention of cardiovascular disease in Switzerland: Methods and results of the National Research Program 1A (NRP1A). Preventive Medicine 14, 482–491.

Heid, I. M., Küchenhoff, H., Miles, J., Kreienbrock, L., Wichmann, H. E., 2004. Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. Journal of Exposure Analysis and Environmental Epidemiology 14, 365–377.

Held, L., 2004. Simultaneous posterior probability statements from Monte Carlo output. Journal of Computational and Graphical Statistics 13, 20–35.

Held, L., Sauter, R., 2016. Adaptive prior weighting in generalized regression models. Biometrics.Doi: 10.1111/biom.12541.

Hu, P., Tsiatis, A. A., Davidian, M., 1998. Estimating the parameters in the Cox model when covariate variables are measured with error. Biometrics 54, 1407–1419.

Hughes, M. D., 1993. Regression Dilution in the Proportional Hazards Model. Biometrics 49, 1056–1066.

Kalbfleisch, J. D., Prentice, R. L., 2002. The Statistical Analysis of Failure-Time Data, 2nd Edition. Wiley, New York.

Kerber, R. A., Till, J. E., Simon, S. L., Lyon, J. L., Thomas, D. C., Preston-Martin, S., Rallison, M. L., Lloyd, R. D., Stevens, W., 1993. A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. Journal of the American Medical Association 270, 2076–2082.

Khan, T. V., Khan, S. S., Akhondi, A., Khan, T. W., 2007. White coat hypertension: Relevance to clinical and emergency medical services personnel. Medscape General Medicine 9, 52.

Klesges, R., Debon, M., Ray, J. W., 1995. Are self-reports of smoking rate biased? Evidence from the second National Health and Nutrition Examination Survey. Journal of Clinical Epidemiology 48, 1225–1233.

Küchenhoff, H., Bender, R., Langner, I., 2007. Effect of Berkson measurement error on parameter estimates in Cox regression models. Lifetime Data Analysis 13, 261–272.

Küchenhoff, H., Mwalili, S. M., Lesaffre, E., 2006. A general method for dealing with misclassification in regression: The Misclassification SIMEX. Biometrics 62, 85–96.

Li, Y., Clyde, M. A., October 2016. Mixtures of g-priors in generalized linear models. Tech. rep., Clemson/Duke University, arXiv:1503.06913v2.

Lindley, D. V., El-Sayyad, G. M., 1968. The Bayesian estimation of a linear functional relationships. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 30, 190–202.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N., 2009. The BUGS project: Evolution, critique, and future directions. Statistics in Medicine 28, 3049–3067.

Martino, S., Akerkar, R., Rue, H., 2011. Approximate Bayesian inference for survival models. Scandinavian Journal of Statistics 38, 514–528.

Martins, T. G., Rue, H., 2014. Extending INLA to a class of near-Gaussian latent models. Scandinavian Journal of Statistics 41, 893–912.

Martins, T. G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA: new features. Computational Statistics & Data Analysis 67, 68–83.

McCrink, L., 2016. Robust joint modelling: The impact of longitudinal outliers. Tech. rep., Unversity of Belfast, UK, submitted.

Muff, S., Keller, L. F., 2015. Reverse attenuation in interaction terms due to covariate error. Biometrical Journal 6, 1068–1083.

Muff, S., Riebler, A., Held, L., Rue, H., Saner, P., 2015. Bayesian analysis of measurement error models using integrated nested Laplace approximations. Journal of the Royal Statistical Society. Series C (Applied Statistics) 64, 231–252.

Myers, R. J., 1954. Accuracy of Age Reporting in the 1950 United States Census. Journal of the American Statistical Association 49, 826–831.

Nakamura, T., 1990. Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models. Biometrika 77, 127–137.

Nakamura, T., 1992. Proportional hazards model with covariates subject to measurement error. Biometrics 48, 829–838.

Plummer, M., 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing.

Prentice, R. L., 1982. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 69, 331–342.

Prentice, R. L., Kalbfleisch, J. D., 1979. Hazard rate models with covariates. Biometrics 35, 25–39.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Richardson, S., Gilks, W. R., 1993. Conditional independence models for epidemiological studies with covariate measurement error. Statistics in Medicine 12, 1703–1722.

Rudemo, M., Ruppert, D., Streibig, J. C., 1989. Random-effect models in nonlinear regression with applications to bioassay. Biometrics 45, 349–362.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society. Series B (Statistical Methodology) 71, 319–392.

Sabanés Bové, D., Held, L., 2011. Bayesian fractional polynomials. Statistics and Computing 21, 309 – 324.

Schafer, D. W., 1987. Covariate measurement error in generalized linear models. Biometrika 74, 385–391.

Schafer, D. W., 1993. Likelihood analysis for probit regression with measurement errors. Biometrika 80, 899–904.

Simon, S. L., Till, J. E., Lloyd, R. D., Kerber, R. L., Thomas, D. C., Preston-Martin, S., Lyon, J. L., Stevens, W., 1995. The Utah leukemia case-control study: dosimetry methodology and results. Health Physics 68, 460–471.

Song, X., Davidian, M., Tsiatis, A. A., 2002. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3, 511–528.

Stefanski, L. A., 1989. Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. Communication in Statistics, Series A 18, 4335–4358.

Stephens, D. A., Dellaportas, P., 1992. Bayesian analysis of generalised linear models with covariate measurement error. In: Bernardo, J. M., Berger, J. O., Dawid, A. P., Smith, A. F. M. (Eds.), Bayesian Statistics 4. Oxford Univ Press.

Tadesse, M. G., Ibrahim, J. G., Gentleman, R., Chiaretti, S., Ritz, J., Foa, R., 2005. Bayesian error-in-variable survival model for the analysis of genechip arrays. Biometrics 61, 488–497.

Therneau, T., 2015. A Package for Survival Analysis in S. Version 2.38.
URL `http://CRAN.R-project.org/package=survival`

Volinsky, C. T., Raftery, A. E., 2000. Bayesian information criterion for censored survival models. Biometrics 56, 256–262.

Von Gunten, E., Braun, J., Bopp, M., Keller, U., Faeh, D., 2013. J-shaped association between plasma glucose concentration and cardiovascular disease mortality over a follow-up of 32 years. Preventive Medicine 57, 623–628.

Wang, H., Heitjan, D. F., 2008. Modeling heaping in self-reported cigarette counts. Statistics in Medicine 27, 3789–3804.

Wang, H., Shiffman, S., Griffith, S. D., Heitjan, D. F., 2012. Truth and memory: Linking instantaneous and retrospective self-reported cigarette consumption. Annals of Applied Statistics 6, 1689–1706.

Wang, X., George, E. I., 2007. Adaptive Bayesian criteria in variable selection for generalized linear models. Statistica Sinica 17 (2), 667–690.

Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In: Goel, P. K., Zellner, A. (Eds.), Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. Vol. 6 of Studies in Bayesian Econometrics and Statistics. North-Holland, Amsterdam, pp. 233–243.

Zelnik, M., 1961. Age heaping in the United States census: 1880-1950. The Milbank Memorial Fund Quarterly 39, 540–573.